REMARKS

Applicants would like to thank Examiner Steadman for the helpful discussion of the issues with the undersigned and Dr. Derbyshire in the telephone conference on October 7. During the discussion, prospective amendments to address the various bases for the Examiner's rejections were discussed. The Applicants representative also noted for the Examiner's consideration the features distinguishing the invention from the cited art. In light of this discussion, Applicants have amended the claims in a way which is though to obviate the Examiner's 35 U.S.C. § 112 rejections and to improve the definition of the invention.

Status of the Claims

Claims 1, 6, 7, 9, 12, and 14-17 are currently pending and under examination. Claims 2-5, 8, 10, 11, 13, and 18 have been canceled without prejudice or disclaimer of the subject matter claimed therein. The pending claims are thought to be allowable for the reasons noted below and otherwise of record.

Amendments to the Claims

Claim 1 has been amended to recite the elements of canceled claim 8.

Claim 17 has been amended to recite the elements of canceled claim 18.

Claims 1 and 17 have also been amended to recite that the yeast is glycosylation deficient as compared to a parent strain and that the AFP exhibit ice recrystallization inhibitory activity. Representative support can be found at page 6, lines 6-13 and at page 15, lines 12-19.

The amendments to the claims do not add new matter.

Claim Objections

Claim 13 is objected to for being a substantial duplicate of claim 9. Without acquiescing to the merits of the objection, Applicants have canceled claim 13.

Claims 1, 14, 15, and 17 are objected to for failing to recite what the homology is to. Applicants have amended these claims to recite "amino acid homology", as suggested in the Office Action. It is therefore believed that the claims as amended are in the requested improved form.

Rejection Under 35 U.S.C. § 112, Second Paragraph

Claims 1, 6-9, and 12-18 and 12 are rejected under 35 U.S.C. § 112, second paragraph, as allegedly failing to particularly point out and distinctly claim the subject matter regarded as the invention.

The Office Action alleges that claim 1 fails to provide a standard for determining the requisite degree of protein mannosyl transferase deficiency. Applicants have amended the claims to recite that the deficiency is compared to a parent strain.

The Office Action also alleges that claim 17 is confusing for recitation that the yeast is a protein mannosyl transferase. Applicants have amended the claim to recite that the yeast is a strain deficient in pmt1 and/or pmt2.

The Office Action also alleges that claim 17 is indefinite in recitation of functional equivalents for failing to recite the function the equivalents can achieve. Applicants have amended claim 17 to recite that the type III AFP exhibit ice recrystallization inhibitory activity.

The Office Action also alleges that claim 17 is indefinite for failing to set the metes and bounds of the anti-freeze protein (AFP). Applicants have amended the claims to recite that the nucleic acid encode a type III HPLC-12 AFP. It is therefore respectfully requested that the Examiner's rejections under 35 U.S.C. § 112, second paragraph, be withdrawn.

Rejection Under 35 U.S.C. § 112, First Paragraph

A. Claims 1, 6-9, and 12-18 are rejected under 35 U.S.C. § 112, first paragraph, as allegedly failing to comply with the written description requirement.

The Examiner alleges that the claims do not recite a functional limitation that the type III antifreeze protein must achieve. The claims, as amended, recite that the type III AFP exhibit ice recrystallization inhibitory activity. The portions of AFPs responsible for ice recrystallization inhibitory activity are known in the art (*see, e.g.*, Baardsnes *et al.* J. Biol. Chem 278: 38942-38947, 2003, and Graether *et al.* J. Biol. Chem. 274: 11842-11847, 1999, both attached hereto). Accordingly, it is believed that the claims as amended appropriately refer to a functional limitation for the type III AFPs.

The Office Action also alleges that the specification fails to provide written description

for the genus of yeast deficient in pmt1 and/or pmt2. Without acquiescing to the merits of the rejection, Applicants have amended the claims to recite that the yeast is *S. cerevisiae.* Accordingly, it is believed the claims, as amended, comply with the written description requirement. Applicants therefore respectfully request that the Examiner's rejection under 35 U.S.C. § 112, first paragraph, be withdrawn.

B. Claims 1, 6-9, 12-18 are rejected under 35 U.S.C. § 112, first paragraph, as allegedly failing to enable the claimed invention. The Examiner alleged that the specification fails to enable producing any type III AFP polypeptide using any yeast host deficient in protein glycosylation.

The Examiner alleges that the specification does not enable a method of producing any type III AFP in any type of fungal host cell deficient for glycosylation without a functional limitation. The claims, as amended, recite that the type III AFP exhibit ice recrystallization inhibitory activity and that the yeast is *S. cerevisiae.* Accordingly, it is believed that these amendments overcome the basis for this enablement rejection. It is therefore respectfully requested that the Examiner's rejection under 35 U.S.C. § 112, first paragraph, be withdrawn.

Rejection Under 35 U.S.C. § 103(a)

Claims 1-9, 12, and 13 are rejected under 35 U.S.C. § 103(a) as allegedly being obvious under Chapman *et al.* (WO 97/02343) ("Chapman") in view of Ng *et al.* (U.S. Patent Application Publication 2002/0068325) ("Ng") and Gentzsch *et al.* (FEBS Lett 377: 128-130, 1995) ("Gentzsch").

The Examiner alleges that Chapman discloses a recombinant type III AFP HPLC-12 produced in *Saccharomyces cerevisiae*, Ng discloses O-linked glycosylation in the ER by pmts, and the Gentzsch discloses pmt1 and pmt2 deletion mutants. The Examiner alleged it would be obvious to recombinantly produce the AFP of Chapman in the pmt mutants of Gentzsch given the disclosure concerning glycosylation disclosed by Ng.

The Applicants respectfully submit that there is no basis in the cited references to warrant the combination proposed by the Examiner.

Chapman discloses a production of a type III AFP HPLC-12 produced in *S. cerevisiae.* It

is of significance that the AFP is properly produced in yeast when assessing the obviousness to combine the elements from the references. Ng discloses use of pmt mutants as a means to overcome "aberrant" glycosylation that leads to rapid degradation of the recombinant protein (*see* Ng at paragraph [0071]). The aberrant glycosylation disclosed by Ng prohibits proper folding of the recombinant protein and thus the protein is not secreted and is rapidly degraded. The AFP disclosed by Chapman, however, is reportedly produced efficiently in wild type yeast. Accordingly, one skilled in the art would not determine that the AFP of Chapman needs to be produced in a pmt mutant yeast or that to do so would help in any way. As the Office Action acknowledges, prediction of what glycosylation does to a protein is unpredictable. One skilled in the art would not have an inkling of what altering glycosylation would do to the type III AFP. Indeed, Chapman does not even disclose that the type III AFP is glycosylated at all when produced in wild type yeast.

Furthermore, no cited reference discloses that pmt1 or pmt2 are the pmt proteins responsible for glycosylation of the type III AFP protein in yeast. One skilled in the art would not know from the art what pmt or combination of pmts are responsible for glycosylating type III AFPs or that the indicated glycosylation could be usefully omitted in the preparation of type III AFP.

The Federal Circuit recently stated that the decision of the Supreme Court in *KSR* with regards to obviousness requires that there be "a reason that would have prompted a person of ordinary skill in the relevant field to combine the elements in the way the claimed new invention does." *Takeda Chemical Industries, Ltd. v. Alphapharm Pty., Ltd.* 492 F.3d 1350, 1356-1357 (Fed. Cir. 2008) (citing *KSR International Co. v. Teleflex Inc.* 127 S.Ct. 1727 (2007)). Accordingly, there must be a reason present for a person skilled in the art to combine the references in order to find the invention obvious. As Chapman does not report his AFP to even be glycosylated when produced in yeast and no reference describes what effect glycosylation might have on the product or which pmt enzymes are responsible to glycosylate an AFP, there is no reason present in the cited references that would lead one to make the Examiner's selective combination of the references to reach the invention and to otherwise make it obvious. It is therefore respectfully requested that the § 103(a) rejection be withdrawn.
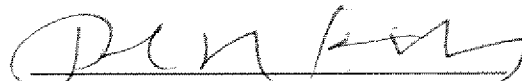
Conclusion

The foregoing amendments and remarks are thought to obviate the basis for the Examiner's rejections and to otherwise place the application in condition for allowance. Accordingly, Applicants respectfully request reconsideration and allowance of the pending claims. Should an interview be helpful to further prosecution of this application, the Examiner is invited to telephone the undersigned.

If there are any additional fees due in connection with the filing of this response, please charge the fees to our Deposit Account No. 50-0310. If a fee is required for an extension of time under 37 C.F.R. §1.136 not accounted for above, such an extension is requested and the fee should also be charged to our Deposit Account.

Dated: **October 20, 2008**
Morgan, Lewis & Bockius LLP
Customer No. **09629**
1111 Pennsylvania Ave., N.W.
Washington, D.C. 20004
202-739-3000

Respectfully submitted
Morgan, Lewis & Bockius LLP

Paul N. Kokulis
Registration No. 16,773

Antifreeze Protein Dimer

WHEN TWO ICE-BINDING FACES ARE BETTER THAN ONE*

Jason Baardsnes‡, Michael J. Kuiper, and Peter L. Davies§

From the Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

A naturally occurring tandem duplication of the 7-kDa type III antifreeze protein from Antarctic eel pout (*Lycodichthys dearborni*) is twice as active as the monomer in depressing the freezing point of a solution. We have investigated the basis for this enhanced activity by producing recombinant analogues of the linked dimer that assess the effects of protein size and the number and area of the ice-binding site(s). The recombinant dimer connected by a peptide linker had twice the activity of the monomer. When one of the two ice-binding sites was inactivated by site-directed mutagenesis, the linked dimer was only 1.2 times more effective than the monomer. When the two monomers were linked through a C-terminal disulfide bond in such a way that their two ice-binding sites were opposite each other and unable to engage the same ice surface simultaneously, the dimer was again only 1.2 times as active as the monomer. We conclude from these analyses that the enhanced activity of the dimer stems from the two ice-binding sites being able to engage to ice at the same time, effectively doubling the area of the ice-binding site.

Fish are protected from freezing by antifreeze proteins (AFPs),[1] which bind to the surface of nucleating ice crystals in their body fluids, thereby reducing their freezing point below that of the ocean (1–3). AFPs create a local curvature of the ice between adsorbed AFPs, which makes it energetically unfavorable for liquid water to join the ice surface (4). This process, known as the Kelvin effect, produces a non-equilibrium reduction of the freezing point of ice below the melting point (5). The difference between the melting and freezing points is the thermal hysteresis gap, and it is within this temperature range that ice growth is prevented and fish are protected from freezing.

Type III AFPs belong to one of several structurally distinct antifreeze protein families found in fishes (6–8). These 7-kDa proteins have a compact β-stranded structure (9–13) that shows homology to the C-terminal domain of sialic acid synthase (14). Structure-function studies have localized the ice-binding residues to a flat, amphipathic surface on the protein (11, 15, 16). This ice-binding face includes several conserved hydrophilic residues (Gln^9, Asn^{14}, Thr^{15}, Thr^{18}, Gln^{44}) that potentially form hydrogen bonds with water molecules on the ice surface (10, 17), flanked by hydrophobic residues (Leu^{10}, Ile^{13}, Leu^{19}, Val^{20}, Val^{41}) on the periphery. Altogether, these residues are thought to make favorable van der Waals contacts with the ice surface (10, 12, 18–20). In addition, there may also be stabilizing entropic effects from bringing this somewhat hydrophobic surface into contact with ice (10, 21). Type III AFP was first reported to bind to the primary prism plane {10-10} of ice (22). However, Antson *et al.* have recently shown it can bind to several planes lying parallel with or at an acute angle to the *c*-axis of the ice crystal (9).

Approximately 20 type III isoforms from five species of zoarcid fishes have now been sequenced, with an overall 50% sequence identity. One isoform (designated RD3) from the Antarctic eel pout (*Lycodichthys dearborni*) has two type III AFPs joined in tandem by a nine-amino acid linker peptide (23). This duplicated AFP with two similar AFP domains has been referred to in the literature as an intramolecular dimer (24) although there is no suggestion that the tandemly repeated AFPs contact each other. NMR analysis of the 14.7-kDa RD3 isoform indicates that the linker region is fairly flexible and may allow both ice-binding faces to engage the ice surface at the same time (24). RD3 was reported to have twice the molar activity of the monomeric isoforms RD1 and RD2 (23) and anywhere from 1.5 to six times the activity of the recombinant N-terminal domain alone depending on the concentration tested (24). Biosynthetic trimers and tetramers of type III AFP show a further modest increase in activity (25).

Here we have investigated the basis for the increased activity of a recombinant type III AFP dimer modeled on the RD3 isoform. The dimer was made using two identical copies of the *Macrozoarces americanus* HPLC-12 isoform derivative (Swiss Protein ID P19614). Currently, this is the best characterized type III AFP isoform. Its 1.15-Å x-ray structure has been solved independently by two groups (9), a high precision NMR structure is available (10), and its ice-binding site has been defined by site-directed mutagenesis (11, 15, 16, 18). As a result of these findings, the activities of numerous mutants and derivatives have been determined as a function of protein concentration. The intramolecular dimer was made by connecting two monomers in a tandem, head-to-tail arrangement by an RD3-type linker sequence. This construct was termed wlwAFP, for wild type-linker-wild type AFP, and its activity was compared with that of the HPLC-12 monomer.

Fusions of AFPs with non-antifreeze proteins can enhance activity simply through an increase in size (26). To quantify this effect, one of the ice-binding sites in the dimer was elimi-

[1] The abbreviations used are: AFP, antifreeze protein; wlwAFP, wild type-linker-wild type AFP; wlxAFP, wild type-linker-knockout AFP; wsswAFP, disulfide-linked wild type AFP homodimer; FPLC, fast performance liquid chromatography; HPLC, high performance liquid chromatography; DTT, dithiothreitol.

A)

	Y	A	A	D	G	T	T	S	P	G	L	K	A	M
5'-GT	TAC	GCT	GCT	GAC	GGT	ACT	ACT	AGC	CCA	GGT	CTG	AAA	GC-3'	
3'-CGA		CGA	CTG	CCA	TGA	TGA	TCG	GGT	CCA	GAC	GAC	TTT	CGA	T-5'

FIG. 1. Synthetic linker oligonucleotides and sequence alignment of RD3, wlwAFP and wlxAFP. *A,* synthetic oligonucleotides used in construction of the linker region. The RD3 linker amino acid sequence is *underlined. B,* a blastp pairwise alignment of wlwAFP and RD3, with single amino acid substitutions for wlxAFP identified below the wlwAFP sequence. wlwAFP and RD3 have 77% amino acid identity.

B)

```
      01        10        20        30        40        50
wlw   MNQASVVANQLIPINTALTLVMMRSEVVTPVGIPAEDIPRLVSMQVNRAVPLGTTLMPDM
      N+ASVVANQLIPINTALTL+MM++EVVTP+GIPAE+IP LV MQVNRAVPLGTTLMPDM
RD3   -NKASVVANQLIPINTALTLIMMKAEVVTPMGIPABEIPNLVGMQVNRAVPLGTTLMPDM

      60        70        80        90        100       110
wlw   VKGYAADGTTSPGLKAMNQASVVANQLIPINTALTLVMMRSEVVTPVGIPAEDIPRLVSM
wlx                                          S   N
       K Y  DGTTSPGLK      SVVANQLIPINTALTLVMM++E V+P GIP+E+I +LV M
RD3   VKNY-EDGTTSPGLK-----SVVANQLIPINTALTLVMMKAEEVSPKGIPSREISKLVGM

      120       130       140
wlw   QVNRAVPLGTTLMPDMVKGYAA
wlx   T
      QVNRAV L  TLMPDMVK Y
RD3   QVNRAVYLDQTLMPDMVKNY--
```

nated by the triple mutation N14S/T18N/Q44T (16) to produce wlxAFP. Thus, a comparison between wlwAFP and the dimer with one ice-binding site knocked out (wlxAFP) could determine how much of the increased activity of the dimer is because of an increase in the overall size of the protein and how much is because of doubling of the area of the ice-binding surface. Finally, to determine whether enhancement in activity results from increasing the number of ice-binding sites, irrespective of the orientation of these faces relative to a planar ice surface, an intermolecular homodimer (wsswAFP) was created through the formation of a disulfide bond between C-terminal Cys residues of an A65C variant. On the basis of modeling, the ice-binding faces of this tail-to-tail dimer are sterically inhibited from simultaneously interacting with the same ice surface.

When the two AFPs were joined in tandem by the 9-amino acid RD3 linker sequence (wlwAFP), their antifreeze activity was twice that of the type III monomer. However, when the two AFPs were joined tail-to-tail by the C-terminal disulfide bond their AFP activity was only 1.2-fold higher than the monomer (wsswAFP). The main difference between these dimers is that the tail-to-head version can potentially engage both ice-binding sites to the same ice surface, whereas in the tail-to-tail dimer the two sites cannot simultaneously dock to a single plane ice. When one of the ice-binding sites of the tail-to-head dimer was inactivated by site-directed mutagenesis (wlxAFP), the net antifreeze activity also decreased to 1.2-fold that of the monomer. Thus, the higher activity of the head-to-tail dimer is attributable in part to its increased size but is mainly because of the flexible linker allowing a doubling of the protein surface in contact with ice.

EXPERIMENTAL PROCEDURES

Construction of Head-to-Tail Dimers wlwAFP and wlxAFP—Clone pT7-7f rQAE contains the synthetic gene for the N-terminally modified HPLC-12 isoform of type III AFP (27). Its modification is the replacement of the signal sequence with an N-terminal methionine residue for expression in *Escherichia coli.* To construct wlwAFP (Fig. 1), pT7-7f rQAE was digested with *Bgl*II and *Bst*EII to release the insert from upstream of the T₇ promoter (*Bgl*II site) to the 3'-end of the antifreeze gene (*Bst*EII site). The *Bgl*II/*Bst*EII fragment was used to encode the N-terminal domain of the recombinant dimer. Clone pT7-7f m1.1, which encodes a C-terminal modification of rQAE (rQAEm1.1) where the YPPA C-terminal sequence is replaced by YAA (27), was digested with *Bgl*II and *Nde*I to remove the upstream promoter region.

The linker sequence was constructed from complementary oligonucleotides and had compatible *Bst*EII and *Nde*I sites at its 5'- and 3'-ends, respectively (Fig. 1). Approximately 50 ng each of the *Bgl*II/*Bst*EI gene insert and the *Nde*I/*Bgl*II gene + vector sequence and 75 pmol of annealed linker were ligated overnight using T₄ DNA ligase

(Invitrogen). The ligated products were transformed into *E. coli* JM-83, and positive clones were selected based on the appearance of a 500-bp *Nde*I/*Pst*I digestion fragment, indicating a successful triple ligation. Positive clones encoding wlwAFP were verified by DNA sequencing (Cortec, Kingston, Ontario, Canada).

To make wlxAFP, the same procedure was followed using an inactive triple mutant (N14S/T18N/Q44T) of the rQAEm1.1 gene (16) to supply the *Nde*I/*Bgl*II gene + vector fragment.

Expression and Purification of Recombinant Dimers—Recombinant intramolecular dimers were expressed and purified following protocols used for the monomers (16) with a few minor changes. To limit proteolysis within the flexible linker region, the protein was refolded for 2 h in an ice bath with the inclusion of 0.1 mM phenylmethylsulfonyl fluoride and 1.0 mM EDTA to both the refolding and dialysis buffers. The protein was purified using S-Sepharose FPLC with an increasing NaCl concentration gradient in 50 mM sodium acetate (pH 3.7). The FPLC-pure protein was dialyzed with 10 mM NH₄HCO₃ and subsequently lyophilized. The purified protein was resuspended in H₂O, and concentrations were verified by amino acid analysis (Alberta Peptide Institute, University of Alberta, Edmonton, Canada). Samples were prepared for thermal hysteresis measurements in 100 mM NH₄HCO₃.

Generation of the wsswAFP Intermolecular Dimer—The A65C mutant of rQAEm1.1 was constructed by primer-directed mutagenesis (28) using the sequence 5'-TAAAGGTTACGCTTGTTAAGAATTCGGAT-3' as described previously (27). The protein was expressed and purified from *E. coli* inclusion bodies following standard protocols except for the addition of 10 mM β-mercaptoethanol to the 50 mM sodium acetate (pH 3.7) used as the refolding, dialysis and FPLC buffers. This allowed the protein to remain monomeric and prevented irreversible aggregation of the protein with the cell lysate. The FPLC-purified A65C was dialyzed into 5 mM NH₄HCO₃ at 4 °C and lyophilized. Approximately 5 mg of AFP was resuspended into 0.5 ml oxidization buffer (10 mM Tris-HCl, 1 mM EDTA, 0.1 M NaCl (pH 8.0)) in a 1-ml Eppendorf tube and incubated at 4 °C for 48 h. No other cysteine residues are present in the AFP; therefore alternate inter- or intramolecular disulfides cannot be formed. After 2 days in the oxidizing buffer, disulfide-linked A65C (wsswAFP) was separated from monomer by C₁₈ semi-preparative HPLC (Vydac, Inc.). Samples were loaded at a flow rate of 3.0 ml/min onto the column pre-equilibrated in 35% acetonitrile:isopropanol (2:1) containing 0.05% trifluoroacetic acid and eluted by a linear acetonitrile:isopropanol (2:1) gradient of 1%/min. The peak containing wsswAFP was identified by 10% SDS-PAGE run in the absence of reducing agents, with standard type III monomer and broad range molecular weight marker (New England Biolabs, Inc.) as standards. Protein samples (~1–5 μg) were reacted with 5 μl of 2.5 M iodoacetate (pH 8.0) to alkylate free cysteinyl residues. Samples were mixed with 5 μl sample loading dye without β-mercaptoethanol to retain the integrity of the disulfide bonds. Protein was visualized by staining with Coomassie Brilliant Blue. The HPLC-purified dimer was lyophilized and resuspended in 100 mM sodium phosphate (pH 3.7). The exact concentration of purified dimer was determined by amino acid analysis (Alberta Peptide Institute).

Reduction of Intermolecular Dimer—Reduction of the disulfide-linked dimer by dithiothreitol (DTT) was done *in situ* to accurately
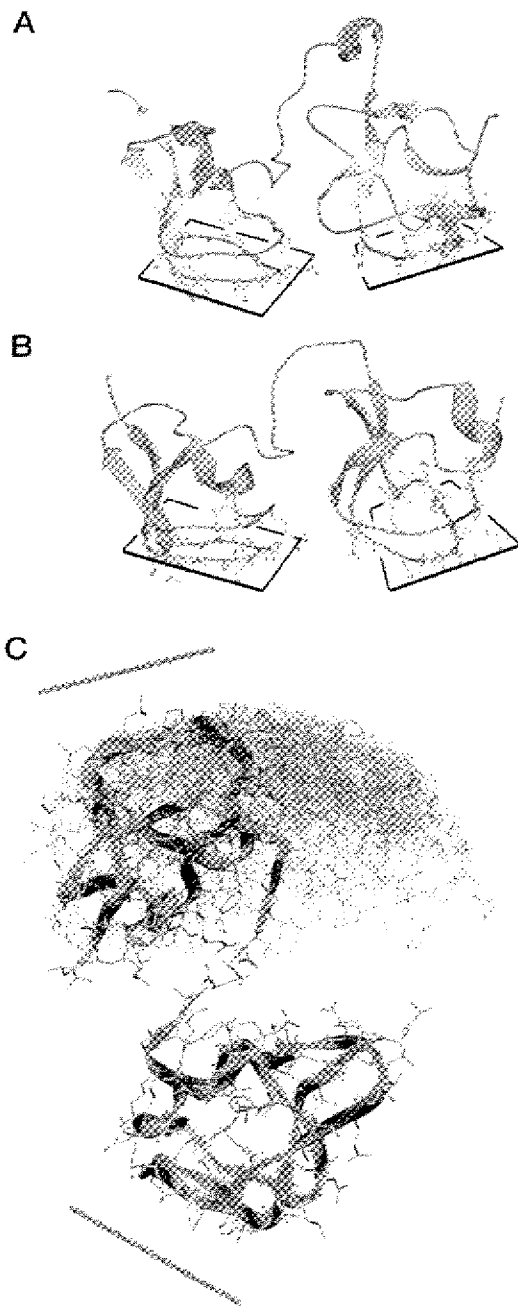
A

B

C

FIG. 2. **Structures of wlwAFP, RD3, and wsswAFP.** *A,* ribbon diagram showing the modeled wlwAFP structure. *B,* the RD3 NMR structure (Protein Data Bank code 1C8A) (24). The side chains of the ice-binding face are shown and are delineated by *black outlines.* The figure was generated by Swiss-PdbViewer (version 3.6) and rendered using POV-Ray for Windows (version 3.1). *C,* WsswAFP model structure generated by SYBYL. The most favorable conformation is shown as a *blue ribbon,* and the *solid lines* indicate the ice-binding faces. The overlap of ice-binding face residues Gln^9 to Val^{20} of the remaining energetically favorable 120 conformations is shown in *green.*

assess the antifreeze activity of the disulfide-linked dimer compared with reduced monomer from the same stock sample. Stock wsswAFP solutions were adjusted to 200 mM NH_4HCO_3 with (reduced) or without (oxidized) 50 mM DTT. Reduction was achieved at 4 °C overnight. The oxidized and reduced wsswAFP samples were assayed for thermal hysteresis activity, and their integrity was monitored by SDS-PAGE.

Thermal Hysteresis Assays and Ice Morphology Determination— Thermal hysteresis is defined as the temperature difference (°C) between the melting point and non-equilibrium freezing point of an AFP solution. Mutant and wild type AFPs were assayed for thermal hyster-
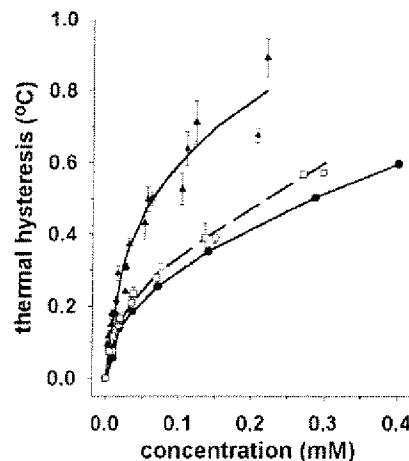

FIG. 3. **Thermal hysteresis activity curves of wlwAFP and wlx-AFP.** Thermal hysteresis activity curves of wlwAFP (*triangles*), wlx-AFP (*squares*), and rQAEm1.1 (*circles*). Standard deviations of wlwAFP and wlxAFP are shown as *vertical bars.* Each point on the activity curve was assayed in triplicate. The fully active wlwAFP curve represents the best fit from three independent trials, and the wlxAFP curve is based on two independent trials.

esis activity as described previously (29). Ice crystals were observed using a Leitz 22 microscope, and ice gap growth of more than 0.2 μm/s signifies that the solution freezing point has been reached or exceeded.

*Molecular Modeling of wlwAFP and wsswAFP—*The initial wlwAFP model was constructed by submitting the entire sequence to Swiss-MODEL (30) using the RD3 NMR structure (24) (Protein Data Bank code 1C8A) as a template. Compared with RD3, wlwAFP has an additional amino acid (Ala) at the C-terminal end of the dimer linker and four residues (MNQA) at the N terminus of the second AFP of the dimer. (Fig. 1). This segment generated an unrealistic loop in the Swiss-MODEL structure. However, PredictProtein (Department of Biochemistry and Biophysics, Columbia University) predicted a small α-helix within this region. Therefore, the Swiss-MODEL structure was manually refined using SYBYL (version 6.5, Tripos Associates, St. Louis, MO) to include this potential helical structure. The structure was then minimized (<0.05 kcal/mol gradient terminated) using a Tripos force field, Gasteiger/Marsilli charges, and distant dependent dielectric. Side chain conformations were verified using PROCHECK (30) (Fig. 2A).

The wsswAFP model was constructed using the molecular modeling software SYBYL, by replacing Ala^{65} to Cys in two type III AFP models (Protein Data Bank code 1MSI) and then creating a disulfide bond between them using the Biopolymer module. Rotational freedom about the disulfide bond of the new dimer was explored by using a short SYBYL Programming Language (SPL) script, which systematically performed every torsion angle combination at 10° increments about the central disulfide bond and the two adjacent sulfur–carbon bonds. Each combination was minimized briefly (250 iterations, Tripos force field, Gasteiger/Marsilli charges, distant dielectric constant), and final energy values were recorded. The ice-binding faces of the 120 most energetically favorable conformations were superimposed to indicate the possible orientations of the two ice-binding faces with respect to one another (Fig. 2C).

RESULTS

*wlwAFP Is Twice as Active as Monomeric Type III AFP—*The recombinant dimers that were overexpressed in *E. coli* predominantly partitioned to the insoluble cell pellet (not shown). The refolded dimer proteins from this fraction eluted from ion-exchange FPLC as sharp peaks at the same NaCl concentration as the monomer (27). This is consistent with the pI values for wlwAFP/wlxAFP and the AFP monomer being so similar (6.1 and 6.0, respectively). SDS-PAGE showed that the purified dimers had the expected 14-kDa mass (not shown).

Thermal hysteresis activity for wlwAFP was 0.53 °C at 0.07 mM compared with 0.26 °C for the monomer (Fig. 3). Across the concentration range of 0 to 0.20 mM, the wlwAFP dimer gave a
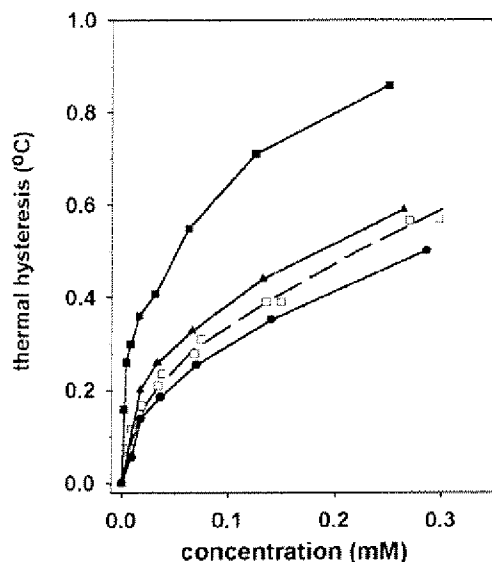
FIG. 4. **Thermal hysteresis activity of type III AFP fusion proteins.** Activity of maltose-binding protein (50 kDa, *closed squares*) and thioredoxin fusion protein (20 kDa, *closed triangles*) (data from Deluca *et al.* (26)), and wlxAFP (14 kDa, *open squares*). Activity of the m1.1 monomer is shown as *closed circles*.
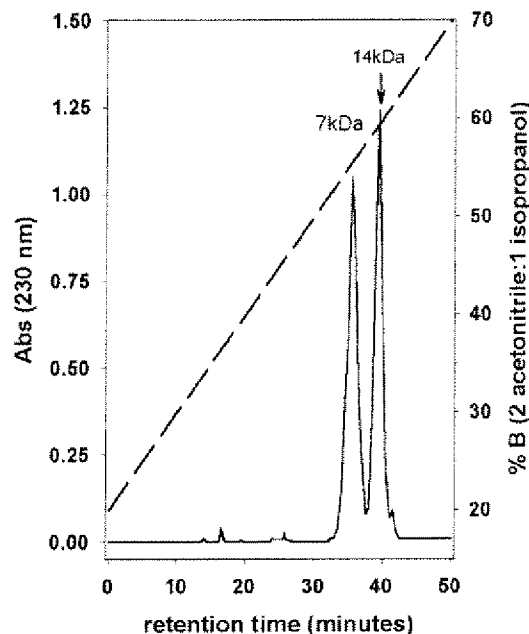
FIG. 5. C_{18} **semipreparative HPLC separation of oxidized wss-wAFP dimer (14 kDa) from A65C monomer (7 kDa).** Gradient of solvent B is shown by the *dotted line*.

2-fold enhancement in activity compared with the monomer. This augmentation is in close agreement with the original findings of Wang *et al.* (23), where the native RD3 isoform at 0.7 mM increased activity by ~1.8-fold on a molar basis compared with the RD1 and RD2 monomer isoforms. This level of enhancement was also reported by Miura *et al.* (24) for recombinant RD3 at a concentration of 0.5 mM when compared with the novel RD3-N1 monomer.

In the wlxAFP dimer, the ice-binding site in the C-terminal domain has been eliminated by the triple mutation N14S/T18N/Q44T. This loss of one of the two ice-binding sites consistently reduced the activity of the dimer by ~40% compared with wlwAFP across the concentration range studied (0 to 0.20 mM) (Fig. 3). However, the wlxAFP dimer levels of activity were still higher than the monomer. The small but constant 20% increase in activity of wlxAFP compared with monomer is consistent with data previously reported by DeLuca *et al.* (26), where increasing the size of type III AFP by fusion to other proteins increased its AFP activity (Fig. 4). Fusion proteins of 50 kDa (type III AFP with a maltose-binding protein fusion), and 20 kDa (type III AFP with a thioredoxin fusion) generated thermal hysteresis activity values of 0.60 and 0.36 °C, respectively, at 0.07 mM. The wlxAFP, at 14 kDa, follows this trend with a thermal hysteresis value of 0.31 °C at 0.07 mM. The activity curve of this protein falls between the type III monomer and 20-kDa thioredoxin-AFP fusion protein across the concentration range tested (0 to 0.25 mM).

Tail-to-Tail Disulfide-linked Dimer—During separation of the disulfide-bonded dimer of A65C (wsswAFP) from the monomer on C_{18} semi-preparative reversed-phase HPLC, two sharp peaks of approximately equal area eluted with retention times of 35 and 40 min (Fig. 5). The proteins in the first and second peaks were identified using nonreducing SDS-PAGE as the monomer and dimer, respectively. There was no cross-contamination of the peaks nor any change in disulfide bonding after peak separation. The order of elution by HPLC was as expected, with the larger dimer protein binding more tightly to the C_{18} column. Based on the area of the HPLC peaks, ~50% of the starting material (2.5 mg) formed a dimer.

A portion of the purified wsswAFP was reduced by the addition of DTT. The complete reduction of the newly formed disul-

fide bond was possible when an alkaline pH was maintained with NH_4HCO_3 (data not shown). This allowed parallel analyses of the same sample in both the monomer and dimer states with absolute confidence in the relative protein concentrations (Fig. 6). After reduction, the molar concentration of the reduced A65C monomer was exactly twice that of the starting wsswAFP. Thermal hysteresis activity measurements showed that the oxidized wsswAFP dimer had 20% more activity than the reduced A65C monomer at protein concentrations of 0.02, 0.07, and 0.2 mM (Fig. 7). This activity augmentation is much less than that seen with the wlwAFP head-to-tail dimer but is similar to the activity enhancement seen for the wlxAFP dimer with one inactive ice-binding surface.

The Conformation of the wsswAFP Dimer—In the absence of a structure for the wsswAFP, the relative orientation of the two ice-binding sites was established by modeling. Both AFP domains were systematically rotated by 10° increments around the disulfide bond and adjoining carbon–sulfur bonds followed by energy minimization at each interval. This allowed for a large sampling of the total possible conformers that could be adopted by the two AFP domains about the disulfide bond. The central Ala[16] residues of each ice-binding face are placed nearly 40 Å apart in the most energetically favorable dimer structure (shown in blue in Fig. 2C). The superimposed trace of the 120 most energetically favorable conformations reveals that both ice-binding faces will occupy opposite surfaces of the dimer and are incapable of simultaneously engaging the same ice surface.

DISCUSSION

Activity of wlwAFP Is Analogous to That of RD3—The recombinant dimer wlwAFP shows an approximately 2-fold increase in antifreeze activity as compared with the monomer. This activity increase is equivalent to the value for the natural dimer reported by Wang *et al.* (23) and is within the range of values determined by Miura *et al.* (24) for recombinantly produced RD3. By confirming this enhancement using a *de novo* designed dimer made from identical monomers that are structurally and functionally well characterized, we show that this property is not a peculiarity of RD3. It can be reconstituted
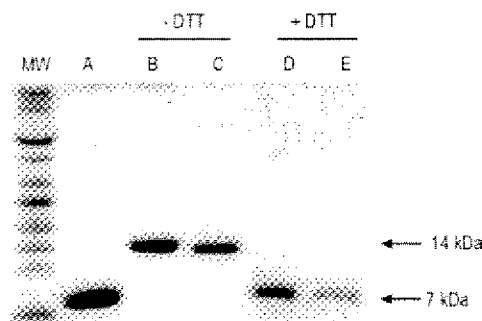
FIG. 6. **SDS-PAGE analysis of wsswAFP in the presence of DTT.** Reduction of wsswAFP with DTT for thermal hysteresis activity readings. *Lane A*, type III AFP QAEm1.1 control; *lane B*, 0.24 mM HPLC purified wsswAFP; *lane C*, 0.38 mM wsswAFP containing 200 mM NH$_4$HCO$_3$ (without DTT); *lane D*, 0.38 mM A65C containing 200 mM NH$_4$HCO$_3$ and 50 mM DTT; *lane E*, 0.048 mM A65C containing 200 mM NH$_4$HCO$_3$ and 50 mM DTT.

using AFPs that differ slightly in sequence from the AFPs present in the RD3 dimer. The doubling of antifreeze activity for a protein of the same molarity has potential applications in biotechnology (31, 32) and is of mechanistic interest. We have considered and investigated several explanations for this phenomenon. Specifically we have examined the effects of the size number of ice-binding sites and the total area of ice-binding site in contact with ice.

Activity Enhancement of wlxAFP Is Based on Size—As predicted by previous observations (26), there was a small gain in activity for the recombinant dimer containing an inactive domain (wlxAFP). Inactivation of the second domain was based on the type III rQAEm1.1 variant, N14S/T18N/Q44T, which was shown previously to be completely inactive from 0 to 0.7 mM (16). Therefore, the 20% gain in activity could not be attributed to antifreeze activity in the C-terminal domain. This activity increase does, however, fit well with the data showing enhanced activity based on size increases using N-terminal fusion proteins (26). The fact that the inactive partner in the wlxAFP fusion is on the C-terminal side does not seem to affect the result. The wlxAFP activity curve fits between the activity curves of the monomer (7 kDa) and thioredoxin fusion (20 kDa). Indeed, there is such a good correlation between fusion protein mass and activity that it is possible to confirm the mass of wlxAFP by interpolation of the data from the activity curves of the monomer and the 20- and 50-kDa fusion proteins. Overall, it is clear that only a small component (~20%) of the dimer activity is due to the increased size of the AFP.

Two Ice-binding Sites Are Not Necessarily Better than One—We considered the possibility that the enhanced activity of the type III AFP dimers might be due in part to two ice-binding sites increasing the probability of binding to ice. The A65C intermolecular dimer was constructed to place the two ice-binding sites at opposite ends of the protein. The cysteine residue is at the C terminus of the monomer, and an intermolecular disulfide bond should form a dimer that is constrained from having both AFP domains simultaneously engage the ice surface. Molecular modeling confirmed that none of the energetically favorable structures would allow simultaneous engagement of both sites to a flat ice surface. The 20% increase in activity of the A65C dimer is comparable with that obtained with wlxAFP and can be accounted for by the increase in size of the AFP. The A65C dimer essentially serves as a fusion protein, and there is no advantage to the extra ice-binding site if it cannot be bound simultaneously to ice. It is important to note that reduction of the disulfide bond allowed for a built-in internal control between the monomer and dimer concentrations. This is especially important at low AFP concentrations where
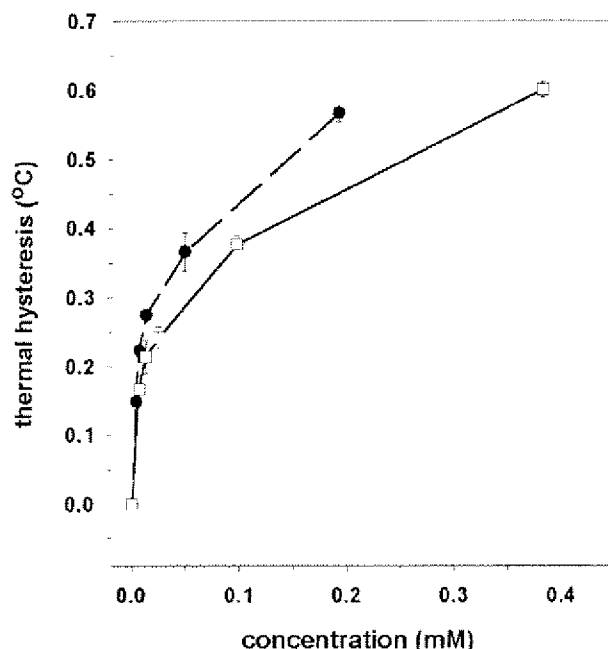

FIG. 7. **Thermal hysteresis activity of wsswAFP.** Thermal hysteresis activity curves of A65C + DTT (*squares*) and wsswAFP − DTT (*circles*).

small changes in concentration have a large effect on antifreeze activity, and accurate concentrations are essential for a meaningful comparison between AFP molecules with similar activity.

Ice-binding Site Area Is the Key to Increased Antifreeze Activity—The wlwAFP dimer has the same mass and number of ice-binding sites as the A65C dimer, wsswAFP, but is considerably more active. The critical difference is the head-to-tail arrangement of the two domains and their flexible linker, which permits the two AFPs to bind the ice surface simultaneously. On the basis of the solution structure of RD3, Miura *et al.* (24) determined that only a small fraction of the total RD3 conformers would be oriented to engage both ice-binding sites simultaneously to the {1 0-1 0} primary prism plane ice surface. This is based on the observation that only one of the NMR solution structures is correctly aligned to accomplish this. However, Antson *et al.* (9) have recently reported that type III AFP can bind to a set of similar crystallographic planes. This includes secondary prism planes ({2–1-1 0}) and primary pyramidal planes ({2 0–2 1}) in addition to the primary prism plane. Therefore, a larger array of potential ice-binding sites exists for RD3 than previously identified, and this will increase the number of active AFP conformations in solution. The second ice-binding face will not be required to bind to the exact same ice plane as the first, allowing for greater flexibility in ice-binding site selection. Verification of the modeled wlwAFP structure using PROCHECK revealed that both ice-binding faces of the derived structure have the potential to engage a flat ice surface simultaneously while satisfying all geometric constraints.

AFP Activity as a Function of Ice-binding Site Area—There is increasing evidence now that antifreeze activity can be enhanced by increasing the length/area of the ice-binding site. The type I AFP9 isoform from winter flounder serum at 52 amino acids in length (4.3 kDa) has an extra 11-amino acid repeat compared with the common serum isoforms HPLC-6 and -8 and therefore has a larger ice-binding face (33). It demonstrates higher antifreeze activity, almost twice that of the more commonly studied 37-amino acid HPLC-6 isoform (3.3 kDa) (34). The same trend is also observed with antifreeze

glycoproteins from Antarctic cod (35). These glycoproteins comprise Ala-Ala-Thr repeating units with an *O*-linked disaccharide on the Thr residue. On a weight basis, AFGP-5, with a molecular weight of 10,500, has approximately twice the activity of AFGP-7 and -8, both of which have an approximate molecular weight of 3000. Another example of increasing antifreeze activity with increasing ice-binding site area is the spruce budworm AFP. A 12-kDa isoform (501) with two additional coils of β-helix is more than twice as active as the shorter 9-kDa isoform (36). As the disulfide-bonded A65C dimer has demonstrated, there is little activity gained from simple dimerization of the AFP domains. Maximal increase in activity requires the ability to increase contact with the ice surface. Because type I AFPs, AFGPs, and spruce budworm AFP are highly repetitive linear proteins, simply lengthening the protein will increase the size of the ice-binding face contacting the ice surface. However, because the type III AFPs are globular proteins with a well defined ice-binding face, the surface area cannot be increased simply by increasing the size of the monomer. Dimerization via the flexible linker is an efficient way to accomplish this task.

As mentioned previously, antifreeze proteins are thought to bind to the ice surface, forcing the ice to grow into a convex surface between bound AFPs. The enhancement seen with fusion proteins is thought to be due to the larger protein reducing the radius between bound molecules, which in turn makes it more difficult for water to join the ice lattice (26). By the same token, AFPs with larger ice-binding sites (*i.e.* a larger "footprint" on the ice) will decrease the area between bound proteins, thereby reducing the radius of the curved ice fronts. Recombinant dimers based on the type III RD3 isoform have shown that activity enhancement arises primarily from an increase in the size of the ice-binding surface in contact with ice, with just a small component attributed to the increase in the size of the protein.

REFERENCES

1. Raymond, J. A., and DeVries, A. L. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2589–2593
2. DeVries, A. L. (1983) *Annu. Rev. Physiol.* **45**, 245–260
3. Yeh, Y., and Feeney, R. E. (1996) *Chem. Rev.* **96**, 601–618
4. Knight, C. A., Cheng, C. C., and DeVries, A. L. (1991) *Biophys. J.* **59**, 409–418
5. Wilson, P. W. (1993) *Cryo Lett.* **14**, 31–36
6. Davies, P. L., and Sykes, B. D. (1997) *Curr. Opin. Struct. Biol.* **7**, 828–834
7. Ewart, K. V., Lin, Q., and Hew, C. L. (1999) *Cell. Mol. Life. Sci.* **55**, 271–283
8. Fletcher, G. L., Hew, C. L., and Davies, P. L. (2001) *Annu. Rev. Physiol.* **63**, 359–390
9. Antson, A. A., Smith, D. J., Roper, D. I., Lewis, S., Caves, L. S., Verma, C. S., Buckley, S. L., Lillford, P. J., and Hubbard, R. E. (2001) *J. Mol. Biol.* **305**, 875–889
10. Sonnichsen, F. D., DeLuca, C. I., Davies, P. L., and Sykes, B. D. (1996) *Structure* **4**, 1325–1337
11. Jia, Z., DeLuca, C. I., Chao, H., and Davies, P. L. (1996) *Nature* **384**, 285–288
12. Yang, D. S., Hon, W. C., Bubanko, S., Xue, Y., Seetharaman, J., Hew, C. L., and Sicheri, F. (1998) *Biophys. J.* **74**, 2142–2151
13. Ko, T. P., Robinson, H., Gao, Y. G., Cheng, C. H., DeVries, A. L., and Wang, A. H. (2003) *Biophys. J.* **84**, 1228–1237
14. Sambrook, J., Fritsch, E. F., Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
15. DeLuca, C. I., Chao, H., Sonnichsen, F. D., Sykes, B. D., and Davies, P. L. (1996) *Biophys. J.* **71**, 2346–2355
16. Chao, H., Sonnichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci.* **3**, 1760–1769
17. Jia, Z., DeLuca, C. I., and Davies, P. L. (1995) *Protein Sci.* **4**, 1236–1238
18. Baardsnes, J., and Davies, P. L. (2002) *Biochim. Biophys. Acta* **1601**, 49–54
19. Graether, S. P., DeLuca, C. I., Baardsnes, J., Hill, G. A., Davies, P. L., and Jia, Z. (1999) *J. Biol. Chem.* **274**, 11842–11847
20. Chen, G., and Jia, Z. (1999) *Biophys. J.* **77**, 1602–1608
21. Harding, M. M., Ward, L. G., and Haymet, A. D. (1999) *Eur. J. Biochem.* **264**, 653–665
22. Cheng, C. C., and DeVries, A. L. (1991) in *Life Under Extreme Conditions* (di Prisco, G., ed) pp. 1–14, Springer-Verlag, Berlin
23. Wang, X., DeVries, A. L., and Cheng, C. H. (1995) *Biochim. Biophys. Acta* **1247**, 163–172
24. Miura, K., Ohgiya, S., Hoshino, T., Nemoto, N., Suetake, T., Miura, A., Spyracopoulos, L., Kondo, H., and Tsuda, S. (2001) *J. Biol. Chem.* **276**, 1304–1310
25. Nishimiya, Y., Ohgiya, S., and Tsuda, S. (2003) *J. Biol. Chem.* **278**, 32307–32312
26. DeLuca, C. I., Comley, R., and Davies, P. L. (1998) *Biophys. J.* **74**, 1502–1508
27. Chao, H., Davies, P. L., Sykes, B. D., and Sonnichsen, F. D. (1993) *Protein Sci.* **2**, 1411–1428
28. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
29. Chakrabartty, A., and Hew, C. L. (1991) *Eur. J. Biochem.* **202**, 1057–1063
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
31. Griffith, M., and Ewart, K. V. (1995) *Biotechnol. Adv.* **13**, 375–402
32. Zachariassen, K. E., and Lundheim, R. (1999) in *Biotechnological Applications of Cold-adapted Organisms* (Margesin, R., and Schinner, F., eds) pp. 319–332, Springer-Verlag, Berlin
33. Baardsnes, J., Kondejewski, L. H., Hodges, R. S., Chao, H., Kay, C., and Davies, P. L. (1999) *FEBS Lett.* **463**, 87–91
34. Chao, H., Hodges, R. S., Kay, C. M., Gauthier, S. Y., and Davies, P. L. (1996) *Protein Sci.* **5**, 1150–1156
35. DeVries, A. L. (1982) *Comp. Biochem. Physiol. A* **73**, 627–640
36. Leinala, E. K., Davies, P. L., Doucet, D., Tyshenko, M. G., Walker, V. K., and Jia, Z. (2002) *J. Biol. Chem.* **277**, 33349–33352

Quantitative and Qualitative Analysis of Type III Antifreeze Protein Structure and Function*

Steffen P. Graether, Carl I. DeLuca‡, Jason Baardsnes, Gregory A. Hill, Peter L. Davies, and Zongchao Jia§

From the Department of Biochemistry, Queen's University, Kingston, Ontario, K7L 3N6 Canada

Some cold water marine fishes avoid cellular damage because of freezing by expressing antifreeze proteins (AFPs) that bind to ice and inhibit its growth; one such protein is the globular type III AFP from eel pout. Despite several studies, the mechanism of ice binding remains unclear because of the difficulty in modeling the AFP-ice interaction. To further explore the mechanism, we have determined the x-ray crystallographic structure of 10 type III AFP mutants and combined that information with 7 previously determined structures to mainly analyze specific AFP-ice interactions such as hydrogen bonds. Quantitative assessment of binding was performed using a neural network with properties of the structure as input and predicted antifreeze activity as output. Using the cross-validation method, a correlation coefficient of 0.60 was obtained between measured and predicted activity, indicating successful learning and good predictive power. A large loss in the predictive power of the neural network occurred after properties related to the hydrophobic surface were left out, suggesting that van der Waal's interactions make a significant contribution to ice binding. By combining the analysis of the neural network with antifreeze activity and x-ray crystallographic structures of the mutants, we extend the existing ice-binding model to a two-step process: 1) probing of the surface for the correct ice-binding plane by hydrogen-bonding side chains and 2) attractive van der Waal's interactions between the other residues of the ice-binding surface and the ice, which increases the strength of the protein-ice interaction.

Many poikilothermic organisms have developed antifreeze proteins (AFPs)[1] to resist freezing. Five classes of structurally diverse antifreeze proteins have been found in fish (for a review, see Ref. 1). These proteins act by adsorbing to the surface of ice and increasing the curvature of ice fronts between bound AFPs (2). The freezing point at the surface is depressed,

which in turn inhibits the growth of ice crystals. The difference between the temperature at which the ice begins to grow (burst point) and the temperature at which the ice crystal melts is known as thermal hysteresis (TH) and is used as a measure of AFP activity.

Recently, several structures of type III AFP have been determined (4–6). Based on the high-resolution x-ray structure (4), a model was proposed whereby surface adsorption occurs through a hydrogen bond match between the side chains of Gln-9, Asn-14, Thr-15, Thr-18, Gln-44, and the ice prism plane $\{10\bar{1}0\}$. These polar residues form part of a flat, amphipathic face that is thought to be the ice-binding surface (Fig. 1). However, the significance of the contribution from hydrogen bonds to the AFP-ice interaction has been questioned by several studies since then. A supposedly conservative change of Thr to Ser in type I AFP led to a large loss of TH activity, whereas a change to the hydrophobic residue valine, which is a better space-filling match, caused only a small loss (7, 8). In the study of the high precision NMR structures of type III AFP (5), the authors argue that hydrogen bonds alone are not sufficient to explain the affinity of the protein for ice because formation of hydrogen bonds between ice and solvent water is enthalpically more favorable because of the "perfect" alignment of water with ice. The less favorable interaction between ice and AFP could be overcome by gain in entropy because of the release of protein-associated water into the bulk solvent. Shape complementarity was examined experimentally by mutating Ala-16, a residue that is located in the center of the putative ice-binding surface (9). The loss of activity in these mutants approximately correlates with the size of the residue substituted for Ala. However, structural interpretation of these changes was complicated by shifts in adjacent residues because of the tight packing of residues at the surface.

A fundamental problem in testing binding hypotheses is the difficulty in analyzing interactions in a quantitative manner. In the crystallographic structure determination of the SP-isoform of type III AFP (6), it is also argued that hydrogen bonds are insufficient for tight binding and that flatness is perhaps a more important factor. A flatness search algorithm was designed and used to show that the proposed ice-binding surface is the flattest in the SP-isoform of type III AFP. However, in the case of the QAE-isoform of type III AFP, the putative ice-binding surface is only the second flattest plane,[2] which questions the relative significance of flatness.

In this study, we prepared and expressed an additional 10 mutants of the eel pout type III AFP and determined their structures to probe their contribution to the activity of the protein. Data from previous studies were also re-examined and compared. In addition, we have used a neural network (10, 11) to predict TH activity. Analysis of the structures and the neural
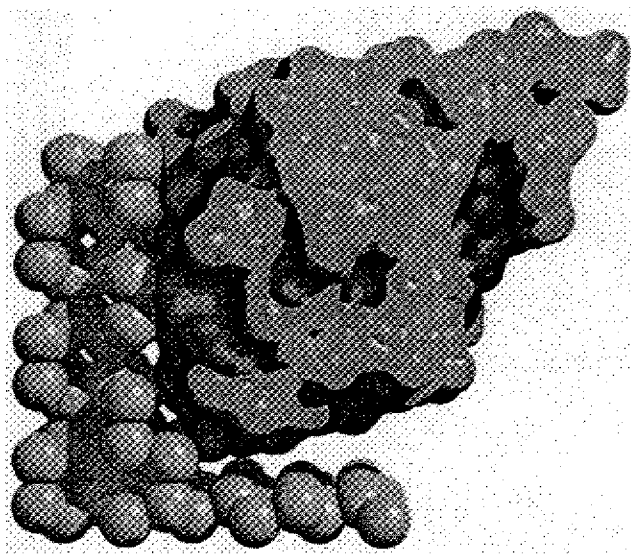

FIG. 1. **Solvent-accessible molecular surface of type III AFP (*red*) docked to a Corey-Pauling-Koltun (CPK) representation of ice (*green*).**

network results shows that changes in van der Waal's interactions and to a lesser extent, hydrogen bonds, are responsible for the loss of activity in type III AFP mutants.

EXPERIMENTAL PROCEDURES

Activity Measurement and X-ray Crystallography of Type III AFP—Mutants of the type III AFP QAE-isoform were made by site-directed mutagenesis as described by Chao *et al.* (12). Thermal hysteresis activity of the mutant proteins was measured and expressed as a percentage of wild-type activity. The mutant proteins crystallized under similar conditions to the wild type proteins (13) with slight variations in ammonium sulfate concentration and pH. Diffraction data (Table I) were collected using a MAR Research imaging plate equipped with a Rigaku rotating anode generator. The data were processed using DENZO/SCALEPACK (14), and structures were refined using X-PLOR (15). The mutated side chain was substituted with Ala in the model in the first round of refinement to prevent bias in the structure determination. In the second round of refinement, Ala was replaced with the mutated residue. The final refinement statistics for the mutant structures are shown in Table 2.

Interpretation of X-ray Structures—Hydrogen bond donors and acceptors involved in ice binding were defined as those oxygen and nitrogen atoms within 2.4–3.5 Å of the modeled ice layer oxygen atoms. The overlap of the mutant protein structures with the native structure was performed using LSQKAB from the CCP4 program (16). Only main-chain atoms were included in the overlap. The positional error of the structures was determined using Luzzati plots (17). Water molecules in the mutant structures were defined as conserved if they were located within 0.45 Å of the water molecule in the wild-type structure, that is, the 0.20-Å positional error from the wild-type protein plus 0.25 Å (mean positional error of the mutants). The figures were generated using Molscript (18) and Raster3D (19).

Neural Network Analysis—As an alternative and quantitative approach to visual analysis, a feed-forward neural network was constructed and tested using the Stuttgart neural network system (SNNS) (20). The SNNS package was chosen for creating and executing the neural network because of its ease of use, availability, and flexibility. Twelve properties (Table III) determined using VADAR (21) of the 16 mutants (9 mutants from this study and 7 previously made ones) and wild-type protein were used as input data, which were scaled such that the lowest value was 0, and the highest was 1. To simplify the analysis and interpretation, only proteins with single mutations were included. Results at the output node represent TH activity scaled from 0 to 1, where 0 is the lowest activity, and 1 is 100% wild-type activity. Twelve input nodes resulted in overtraining, where the neural network was able to predict precisely the TH activity of mutants given in training but not those that were left out. To reduce the number of input nodes, which reduces overtraining, principal component analysis was performed using the NCSS statistics package (22). The first component is a linear

combination of all scaled properties; the second and higher cardinal components cover variance that was not explained in the first component and are orthogonal to all other components. Six components were used to construct a network containing 6 input nodes, 20 hidden nodes, and 1 output node. The optimum number of hidden nodes was empirically determined by varying this number from 8 to 40 in increments of 4. The initial weights of connections between nodes were randomly set to values between −1 and 1, whereas weights after training varied from −10 to 10. Network training was performed for 5000 or more cycles, and default values were used for the remaining parameters. The ability of the network to predict TH activity was tested by cross-validation, which consisted of leaving the TH activity and six components of one mutant out before training. Then the properties of the mutant left out were applied to the trained neural network, and the TH activity was predicted. This was repeated for each of the 17 structures (wild-type protein and 16 mutants). The correlation coefficient established by the "leave-one-out" cross-validation procedure (23) between the predicted and real activity values was calculated using NCSS (22) and used as an indicator of the predictive power of the neural network. To determine which properties were responsible for the predictive ability of the neural network, each group of properties was left out in turn (Table IV), and the cross-validation procedure was repeated. Changes in the correlation coefficient were used as an indicator of the importance of the group of properties in predicting TH activity and, therefore, in the protein-ice interaction.

RESULTS AND DISCUSSION

Validation of X-ray Data—The structure determination and refinement statistics of the 11 mutant structures determined in this study are shown in Tables I and II. With the exception of N- and C-terminal residues and Val-15 in the Thr15Val mutant, there was no ambiguity in side-chain positions. A composite Ramachandran plot of all newly determined structures shows that most non-glycine residues were in the most favored region, a few were in the allowed regions, and none in the disallowed regions (Fig. 2).

Selection of Residues for Mutagenesis—We have combined all the structure information for the QAE-isoform type III AFP mutations previously and newly made (Table V) to provide a comprehensive understanding of how the protein could bind to ice and prevent its growth. To clarify the roles of the various residues, the mutations were based on residues that are proposed to hydrogen bond to ice (Gln-9, Asn-14, Thr-15, Thr-18, and Gln-44), residues that surround these and may interact with the ice through interactions other than hydrogen bonds (Ala-16, Val-20, and Met-21), and those not located on the putative ice-binding face (Ser-24, Pro-29, Pro-33, Glu-35, Ser-42, Arg-39, Asn-46, Arg-47, Asp-58, and Lys-61). The mutations are grouped according to the three regions shown in Fig. 3: (*a*) residues located in the "top" part of the putative ice-binding plane; (*b*) residues located in the "bottom" part of the proposed ice-binding plane; (*c*) residues located along the "bottom" of the protein; (*d*) residues mainly located away from these regions. Although assigning orientations and region boundaries to the protein may be seen as arbitrary, it simplifies the presentation and explanation of the interaction between type III AFP and ice.

Mutation of Residues in Region A—As shown in Table I, mutation of residues in region A resulted in some loss of TH activity, most with TH activities at 70% that of wild-type AFP. Mutation of Thr-18 to Ser resulted in no loss of activity, suggesting that a hydrogen bond between Thr-18O_γ and ice is the important factor in the interaction. For the Thr18Ser mutant, the position of the Ser-18O_γ did not change as compared with that of Thr-18O_γ. However, mutation of Thr-15 to Ser decreased TH activity by 30%, even though Ser-15O_γ would still be able to hydrogen bond to the modeled ice. It is possible that the Ser-15 hydroxyl group, unlike that of Thr-15, is no longer in an ideal position to hydrogen bond to the ice. This is supported by the x-ray structure, where the Ser-15O_γ was translated 1 Å

TABLE I
X-ray reflection data statistics

The data for the wild-type protein has been published previously (4).

Protein	Resolution	Total reflections	Unique reflections	Completeness	R_{sym}
	Å			%	
Wild type	1.25	46,873	16,529	96.6	0.049
Asn14Gln	2.0	22,376	5,329	91.0	0.070
Gln9Thr/Gln44Thr	2.3	13,186	2,599	91.0	0.082
Thr15Ser	1.75	21,851	6,260	98.0	0.079
Thr15Ala	1.7	20,580	6,915	99.4	0.064
Thr15Val	2.3	7,150	2,702	93.6	0.091
Thr18Ser	1.65	40,730	7,566	99.6	0.066
Thr18Ala	2.05	11,371	3,926	96.1	0.079
Val20Ala	1.65	27,744	7,106	93.3	0.065
Met21Ala	2.1	11,961	3,554	94.7	0.075
Arg47His	2.5	7,172	2,100	90.8	0.063

TABLE II
Structure refinement statistics

The data for the wild-type protein has been published previously (4).

Data set	Wild type	Asn14Gln	Gln9Thr/Gln44Thr	Thr15Ser	Thr15Ala	Thr15Val	Thr18Ser	Thr18Ala	Val20Ala	Met21Ala	Arg47His
Unit cell a (Å)	32.33	33.06	32.64	33.28	33.31	32.71	32.72	32.71	33.17	32.59	32.44
b (Å)	38.98	44.07	38.94	39.88	39.90	39.38	39.13	39.37	40.06	39.53	38.71
c (Å)	45.48	44.82	46.38	44.67	46.38	46.51	46.62	46.51	47.45	45.93	47.25
Resolution range (Å)	8.0–1.25	8.0–2.0	8.0–2.3	8.0–1.75	8.0–1.7	8.0–2.3	8.0–1.65	8.0–2.0	8.0–1.65	8.0–2.1	8.0–2.5
R_{factor} (%)	19.8	18.5	18.3	18.3	18.5	16.3	19.0	17.2	18.8	16.6	15.5
(R_{free}, 5%)	22.5	26.1	29.2	26.2	23.0	22.3	21.3	23.8	20.8	22.7	25.1
No. of water molecules	90	48	41	58	46	42	79	59	51	72	42

TABLE III
TH activity measured and predicted by the neural network

TH activity is expressed as percent of wild-type protein at 1 mg/ml. Predicted TH activity was determined with the neural network after the properties for that protein were left out of training.

Protein	Measured TH	Predicted TH
	% wild type	% wild type
Wild type	100	63
Ala16Cys	100	99
Thr18Ser	100	96
Arg47His	100	100
Ala16Met	85	89
Val20Ala	80	92
Met21Ala	80	85
Ala16Thr	75	46
Thr15Ala	70	92
Thr15Ser	70	85
Thr18Ala	70	24
Asn14Gln	67	70
Ala16Arg	60	35
Thr15Val	54	31
Ala16Tyr	33	22
Ala16His	25	77
Thr18Asn	10	50

TABLE IV
Four groups with 12 properties used in the neural network analysis and correlation coefficients after the leave-one-out cross-validation procedure

Neural network training and the cross-validation procedure are described under "Experimental Procedures."

Group	Correlation coefficient after cross-validation with group removed
All properties	0.60
Protein dimensions	0.29
Total volume	
Total ASA	
ASA of side chains	
Hydrophobic	−0.04
Fractional non-polar ASA	
Percent side-chain hydrophobic ASA	
ASA of carbon	
Polar	0.53
ASA of oxygen	
ASA of nitrogen	
Fractional polar ASA	
Charged	0.33
ASA of charged nitrogen	
ASA of charged oxygen	
Fractional charged ASA	

and rotated approximately halfway between the methyl and hydroxyl groups of the wild-type Thr-15 (data not shown). In addition, ice crystals in the presence of Thr15Ser were distinguishable from those of the wild type. They did not grow until just before the burst point, at which time hollow spicules began to form at various points along the crystal. In contrast to the Thr18Ser mutation, the methyl groups of Thr-15 may be required to correctly orient the hydroxyl group so that it is able to effectively hydrogen bond to ice. It would then follow that the 30% loss in activity in the Thr15Ala and Thr18Ala mutants occurred for similar reasons, principally the loss of a hydrogen bond between the protein and ice. The Thr15Val mutation shows that the substitution of the hydroxyl group for a methyl group resulted in approximately half of the TH activity. The electron density of this side chain is poorly defined, unlike the side chains in the other mutant structures. Val-15 may there-

fore not have a well defined position and may exist as one of two main orientations. In one, the side chain can be found in a similar orientation to the Thr-15; in the second orientation, the methyl group pointed toward the hydrophobic core of the protein. When in the latter position, the distance between Val15C$_{\gamma 2}$ and the nearest modeled ice oxygen is 1.85 Å. The resulting steric clash may explain why the presence of valine at this position had a greater effect than the mutation to alanine. It would be of interest to know whether Thr18Val behaves similarly, but several attempts at refolding expressed Thr18Val have failed.
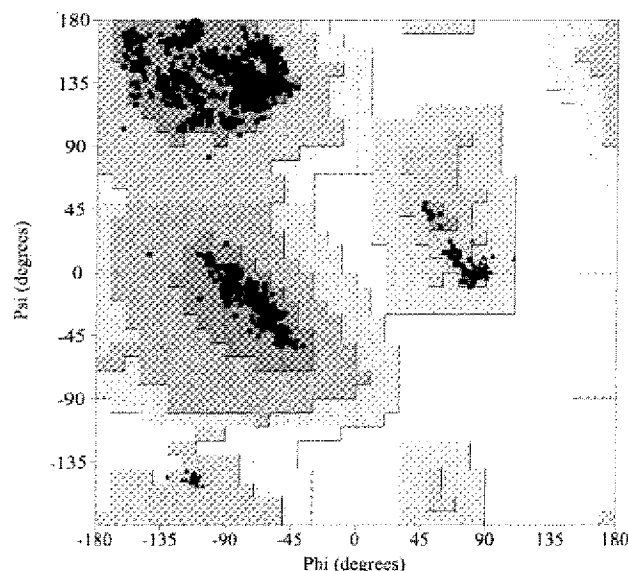
FIG. 2. **Composite Ramachandran plot of type III AFP mutants.** The data were generated using PROCHECK (16). The *dark gray area* represents most allowed regions, whereas the *medium gray areas* represent allowed regions. Glycine residues are represented by *triangles* with other residues represented by *squares*.

TABLE V

Mutation of type III AFP residues and the structure and TH activity

Regions A–D are defined in Fig. 3. Structures were determined as described under "Experimental Procedures." TH activity is expressed as percent of wild-type protein at 1 mg/ml. The structures of wild-type protein, Thr18Asn, and Asn14Ser/Gln44Thr were determined previously (4). ND, not determined.

Region	Protein	Structure resolution	TH
		Å	*% wild type*
	Wild type	1.25	100
A	Gln9Thr	ND	71
	Thr15Ala	1.7	70
	Thr15Ser	1.75	70
	Thr15Val	2.3	54
	Thr18Ala	2.0	70
	Thr18Asn	2.0	10
	Thr18Ser	1.65	100
	Val20Ala	1.65	80
	Met21Ala	2.1	80
B	Asn14Gln	2.0	67
	Asn14Ser	ND	25
	Gln44Thr	ND	47
	Asn14Ser/Gln44Thr	2.3	10
A/B	Gln9Thr/Gln44Thr	2.6	ND
C	Arg47His	2.5	100
	Asp58Ser	ND	100
	Lys61Ile	1.6	100
D	Ser24His	ND	100
	Glu35Lys	1.65	100
	Ser42Gly	1.8	100
	Asn46Ser	1.9	100

The mutant Gln9Thr has approximately the same loss of TH activity as Thr15Ala and Thr18Ala. The structure of Gln9Thr has not been determined, but that of the double mutation Gln9Thr/Gln44Thr has. Because these residues are far apart in the structure, it is assumed that the mutation at Gln-44 does not have an effect on the position of Gln-9. An examination of Gln9Thr in the double mutant showed that Thr-9 is too far away (4.5 Å) from the modeled ice to form a hydrogen bond.

Val-20 and Met-21 are hydrophobic residues located in the proposed ice-binding region. No major changes in the protein structure were seen in either the Val20Ala or Met21Ala mutants. In Met21Ala, there were two minor shifts of Thr-18O$_\gamma$

and Gln9N$_\epsilon$ toward each other, because the methionine side chain no longer separates the two, but these would not drastically affect the ability of Thr-18 and Gln-9 to form hydrogen bonds with the modeled ice surface. These residues appear to contribute to the relative flatness of the proposed ice-binding surface, which could allow favorable weak interactions, such as van der Waal's interactions, to occur. When Val-20 or Met-21 are mutated to Ala, the previously flat surface becomes recessed. The mutations are tolerated fairly well, with only a 20% loss in TH activity. The opposite case, where a substitution causes the side chain to extend above the flat surface, can cause a drastic decrease in TH activity. One example is Thr18Asn, where 90% of the activity was lost (4). The longer side chain may prevent the other ice-binding residues from interacting at the same time with the ice surface. A similar effect is seen with the Ala-16 mutants, where small steric additions to the ice-binding surface by Cys, Met, Thr, or Val resulted in a small decrease in TH activity, whereas bulky groups such as His or Tyr caused a large decrease (9).

Mutation of Residues in Region B—In a previous AFP-ice binding model, it was proposed that Asn-14 and Gln-44 are the first residues to bind to ice (4). Mutation of either of these residues to one with a shorter side chain that still had the ability to form hydrogen bonds, although not necessarily to ice, resulted in a large loss of activity. In the structure of the double mutant Asn14Ser/Gln44Thr, the Ser-14O$_\gamma$ forms a hydrogen bond with Lys61N$_\epsilon$, so that Ser-14 can no longer effectively hydrogen bond to modeled ice. The potential loss of the hydrogen bond at residue Asn-14 is drastic, because the single mutant Asn14Ser has only 25% TH activity. In the case of Asn14Gln, the longer side chain pushes Lys-61 away to make more space so that Gln14O$_\epsilon$ is still able to hydrogen bond with Lys61N$_\epsilon$. However, in the x-ray structure, Gln14N$_\epsilon$ is now too far from the modeled ice to form a hydrogen bond. The decrease in TH activity is not as great as with Asn14Ser (67% *versus* 25% TH activity). Conceivably, Gln-14 may be able to alternate between hydrogen bonding to Lys-61 and to the ice. A mutation at Gln-44 to threonine also results in a structure that is no longer able to hydrogen bond to modeled ice. The activity loss is less severe (50%). The multistep process, where Asn-14 binds first before other hydrogen residues in the ice face, is still a possibility because if other residues bound before Asn-14 did, one might expect mutations of them to cause a similar or more severe decrease in TH activity.

Mutation of Residues in Regions C and D—Mutations in these two regions were made to explore the potential existence of additional ice-binding surfaces on type III AFP. In the case of Pro29Ala and Pro33Ala mutations, the loss of 50% activity was probably the result of changes in the protein backbone conformation because of the structural role often played by proline residues (24). Aside from these two exceptions, mutation of other residues in regions C and D resulted in no detectable loss of TH activity. These mutants are, however, subject to a distinction between those that allowed ice crystal growth during the measurement of TH and those that did not. Residues in region D resulted in no loss of activity and no change in ice crystal morphology. Therefore residues Ser-24, Glu-35, Arg-39, Ser-42, and Asn-46 are not involved, directly or indirectly, in the interaction with ice, and there are no additional ice-binding planes along the top and sides of the protein. In region C, mutation of residues Arg-47, Asp-58, or Lys-61 resulted in a protein with a burst point similar to or slightly higher than that of wild-type protein. Typically, there was slow growth of ice, and there was more variation in TH values. In the case of Lys-61, there could be an interaction between the side chain and the basal plane of ice. The potential formation of additional
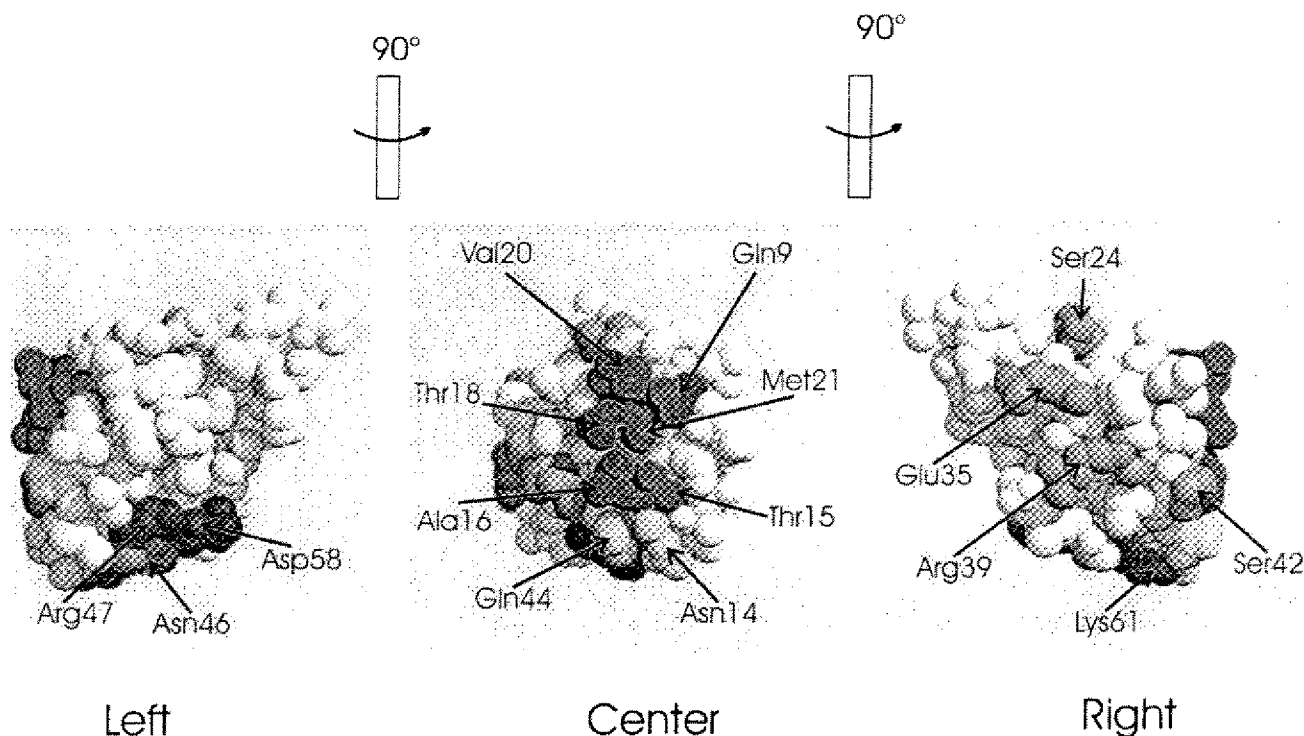
FIG. 3. **Location of mutations in type III AFP; three views of a Corey-Pauling-Koltun (CPK) model of the wild-type III AFP with changed residues marked in color.** *Red*, region A; *yellow*, region B; *blue*, region C; *green*, region D.

hydrogen bonds between the lower protein surface and the basal plane would add to the binding energy, increasing the strength of the interaction. Although residues Arg-47 and Asp-58 are not located on the bottom of AFP, they may indirectly affect the ability of Lys-61 to bind to the basal plane of ice itself or to hydrogen bond to Asn-14. For the latter case, this suggests that residues in region C may not bind to the modeled ice itself, but that losses in activity could be because of the inability of Lys-61 to correctly position Asn-14 for binding to ice.

Neural Network Training—The basic problem of modeling an AFP binding to ice is that no methods that involve the direct detection of interactions between protein and ice (such as solid state NMR) have been reported; thus, the analysis of structure/function relationships have been done in an indirect, qualitative manner. Even more significantly, the modeling of AFP to ice is speculative. Therefore, the TH activities and structures determined in this study and previously were used in a neural network to achieve a quantitative analysis of this interaction that involves proteins but not any AFP-ice model.

Properties were chosen and grouped to account for four possibly important interactions of protein-ice binding. The first group, protein dimensions, consists of total volume, total accessible surface area (ASA), and total ASA of the side chains, where ASA is defined as the area of the protein surface that is in contact with solvent (21). The second group, hydrophobic properties, consists of the nonpolar fractional ASA, percent side-chain ASA, and ASA of carbon. The third group, polar properties, consists of ASA of oxygen, ASA of nitrogen, and fractional ASA of polar atoms, whereas the fourth group, charged properties, consists of ASA of charged nitrogen, ASA of charged oxygen, and fractional ASA of charged groups.

To run the neural network, it is first necessary to optimize several training parameters. Overtraining, that is, the inability of a neural network to predict data it has not seen before, although predicting given data well, increases with the number

of input nodes used. To reduce the number of input nodes, principal component analysis was performed on the 12 properties (25). Six components explained 85% of the variance of all of the properties, whereas fewer or more components resulted in a neural network that could not be trained. Therefore, the six components were used for the six input nodes in the neural network. The size of the hidden layer was varied from 8 to 40 nodes by increments of 4 until the highest correlation factor between experimental and predicted TH activity was determined, which resulted in 20 nodes. The number of training cycles used was 5000 or more cycles.

Neural Network Analysis of Mutants—With the chosen parameters, the average percent difference between the experimentally determined TH activity and predicted TH activity of structures included in training was <0.1%, demonstrating that the network was well optimized.

The correlation between experimental and predicted TH activity for structures used in the leave-one-out cross-validation was 0.60, indicating that the neural network could successfully predict the activity of mutants left out. This value is significant because of the leave-one-out cross-validation procedure (23). Table III shows the measured and predicted activity of wild-type type III AFP and the 16 mutants. Fourteen of the proteins had predicted TH activities that were within 50% of the measured values, with 8 of these being within 25%. Of the remaining 3 mutants, Thr18Ala, Thr18Asn, and Ala16His, the predicted activities did not correspond well with measured activities. It is not clear why these mutants failed, because similar mutations with other residues did not result in a similar failure. Overall, the neural network was well able to predict the activity of a number of diverse mutations of type III AFP. The next step was to remove each group of properties in turn, examine the network for loss of predictive power, and use this as an indicator of which properties are potentially important in type III AFP binding to ice. A simple examination of the neural network to determine the relative weight of each property was not possible

because of the principal component analysis performed and the 20 hidden nodes used.

The changes in the correlation coefficient after leaving each group out in turn before repeating the cross-validation procedure is shown in Table IV. After leaving out protein dimensions and polar or charged properties, the correlation coefficients dropped significantly, but the remaining terms were still able to give some predictive power to the neural network. Thus, these properties could also have a role in ice binding. However, when the hydrophobic group properties were removed, the correlation coefficient essentially dropped to zero. This strongly suggests that the main ability of the neural network to predict the TH activity of the mutants comes from learning about changes in hydrophobic character of the protein surface. The actual interaction could come from attractive van der Waal's interactions and require that type III AFP have a surface that is complementary to that of the ice. This is supported by the data of leaving protein dimensions out, which resulted in the second largest decrease in the correlation coefficient. These properties have no chemical basis but instead are a reflection of changes in the protein shape.

The neural network did not have information of the location of any residue. This would mean, for example, that two mutations that created a protein with identical global surface properties could not be distinguished, even though one may be located in the ice-binding face, but another is far away. Therefore, the neural network would not be able to effectively determine the importance of hydrogen bonds in the interaction. This is not a flaw in the design but was done so as not to bias the neural network with modeled ice. Combining results from structural analysis and neutral network, we envision a mechanism where 1) hydrogen bonds could search for and recognize the correct ice-binding plane and 2) attractive interactions are facilitated by shape complementarity, which allows the protein to remain bound to ice.

Other Factors in Ice Binding—Two additional issues we examined could not be easily integrated into the neural network. Given the large number of mutant structures, the conservation of positioned water molecules was also examined. Many clusters of water molecules were found in the N- and C-terminal region of the protein, which is located away from the putative ice-binding face, whereas few clusters were in the ice-binding region (data not shown). In addition, the structures of three preparations of wild-type proteins have been determined independently, one of which was crystallized at 4 °C (3). None of these structures had water clusters resembling the modeled ice. One may speculate that the water around the ice-binding surface is kept mobile to decrease the energy required to remove them before binding to ice. Such an arrangement may also help to prevent type III AFP acting as an ice seed, because water molecules aligned in an ice-like fashion could promote ice crystal formation before the protein could bind to an existing ice crystal.

Analysis of the structures shows one region where no positioned water molecules are present. Residues in this region are surface-exposed hydrophobic groups (Leu-10, Ile-13, Leu-19, Ile-37, Val-41, and Leu-51) and have no water molecules within 4 Å of any of the atoms in the side chain. They form a hydrophobic ring that circumscribes approximately 2/3 around the protein and is located behind the proposed ice-binding face. The ring is broken by Lys-61, which has a cluster of conserved waters. Ongoing studies of mutation of the residues in this ring to alanine show a dramatic loss of activity.[3] Therefore, part of type III AFP stopping of ice growth may come from the ability of the protein to exclude water from near the ice surface or to keep this water in such a state so that it cannot bind to the ice surface. It is, however, difficult to quantitate the contribution of the hydrophobic ring to the overall interaction and testing of this hypothesis requires further experimentation.

REFERENCES

1. Davies, P. L., and Sykes, B. D. (1997) *Curr. Opin. Struct. Biol.* **7**, 828–834
2. Raymond, J. A., and DeVries, A. L. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 2589–2593
3. Ye, Q., Leinala, E. K., and Jia, Z. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 700–702
4. Jia, Z., DeLuca, C. I., Chao, H., and Davies P. L. (1996) *Nature* **384**, 285–288
5. Soennichsen, F. D., DeLuca, C. I., Davies, P. L., and Sykes, B. D. (1996) *Structure (Lond.)* **4**, 1325–1337
6. Yang, D. S., Hon, W. C., Bubanko, S., Xue, Y., Seetharaman, J., Hew, C. L., and Sicheri, F. (1998) *Biophys. J.* **74**, 2142–2151
7. Chao H., Houston, M. E., Jr., Hodges, R. S., Kay, C. M., Sykes, B. D., Loewen, M. C., Davies, P. L., and Sonnichsen, F. D. (1997) *Biochemistry* **36**, 14652–14660
8. Haymet, A. D., Ward, L. G., Harding, M. M, and Knight, C. A. (1998) *FEBS Lett.* **430**, 301–306
9. DeLuca, C. I., Davies, P. L, Ye, Q., and Jia, Z. (1998) *J. Mol. Biol.* **275**, 515–525
10. Presnell, S. R., and Cohen, F. E. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22**, 283–298
11. Müller, B., Reinhardt, J., and Strickland, M. T. (1995) *Neural Networks. An Introduction*, 2nd Ed., Springer-Verlag New York Inc., New York
12. Chao, H., Soennichsen, F. D., DeLuca, C. I., Sykes, B. D., and Davies, P. L. (1994) *Protein Sci.* **3**, 1760–1769
13. Jia, Z., DeLuca, C. I., and Davies, P. L. (1995) *Protein Sci.* **4**, 1236–1238
14. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
15. Brunger, A. T. (1992) *X-PLOR Manual Version 3.1. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT
16. Collaborative Computing Project No. 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
17. Luzzati, V. (1952) *Acta Cryst.* **5**, 802–810
18. Kraulis, P. J. (1991) *J. Appl. Cryst.* **24**, 946–950
19. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505–524
20. Zell, A., Mache, N., Sommer, T., and Korb, T. (1991) *Informatik-Fachberichte* **285**, 254–263
21. Wishart, D. S., Willard, L., Richards, F. M., and Sykes, B. D. (1994) *VADAR: A Comprehensive Program for Protein Structure Evaluation*, Version 1.2. University of Alberta, Edmonton, Alberta, Canada
22. Hintze, J. (1998) *NCSS Statistical Package*, NCSS Kaysville, Utah
23. Brunger, A. T. (1995) *Nature* **355**, 472–474
24. Chao, H., Davies, P. L., Sykes, B. D., and Sonnichsen, F. D. (1993) *Protein Sci.* **2**, 1411–1428
25. Hotelling, H. (1933) *J. Ed. Psych.* **24**, 417–441, 498–520

[3] J. Baardsnes, and P. L. Davies, unpublished data.